DOCUMENT RESUME

ED 392 827 TM 024 498

AUTHOR Haydel, Juanita B.; And Others
TITLE Assessing Classroom Teacher's Performance Assessments.
PUB DATE Apr 95
NOTE 57p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Classroom Techniques; *Educational Assessment; *Elementary School Teachers; Elementary Secondary Education; Knowledge Base for Teaching; *Outcome Based Education; *Performance Based Assessment; Scoring; *Secondary School Teachers; Teacher Education; *Teacher Made Tests; *Test Construction
IDENTIFIERS Saint Charles Parish Public Schools LA

ABSTRACT

The degree to which performance assessments developed for the classroom demonstrated the principles of quality assessment was explored. Performance assessments were studied in the schools of St. Charles Parish (Louisiana), a district that is implementing outcome based education (OBE). The sample consisted of 92 performance assessments submitted by 79 teachers from elementary, middle, and high schools who were implementing OBE units. An instrument organized around 6 domains and 18 performance criteria was developed for scoring the assessments. Results of the analyses of these tests suggested that teachers' development of classroom performance assessments was not as sound as it could be. Teachers trained in traditional assessment faced real difficulties when asked to assess students using performance assessments. The importance of the study lies in the development of the scoring instrument, which teachers and researchers can use to improve their assessment quality. Appendixes contain an extensive literature review, 22 resources, the assessment samples, the evaluation tool, and a description of the OBE units. (Contains three tables and seven references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 392 827

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JUANITA B. HAYDEL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Assessing Classroom Teacher's Performance Assessments

Juanita B. Haydel and Jeffrey Oescher
University of New Orleans

Mary Banbury
University of New Orleans

Paper presented at the annual meeting of the
American Educational Research Association; San Francisco, CA
April 18-22, 1995

TM024498

BEST COPY AVAILABLE

Assessing Classroom Teacher's Performance Assessments

American education is undergoing severe criticisms on many fronts, and there are efforts to reform or restructure education in response to these criticisms. The challenge of restructuring and redefining schooling has brought with it new challenges for the assessment community. Stiggins (1991) asserts that educators are entering a "whole new era" in terms of assessment and that performance assessment methodology is a central feature of the new era. As educators reform schooling and define achievement targets or outcomes that are more complex, the implications for change in assessment methodology is clear. When outcomes are defined by complex performances or products, traditional assessment methods do not provide an adequate match between the target and purpose of the assessment. A much broader array of assessment is needed, and performance assessments have real potential for measuring many of the valued outcomes. Performance assessments are increasingly being used by today's classroom teachers to match their instructional target with the appropriate assessment method. A review of performance assessment literature is found in Appendix A.

Changes in the mission of schooling, public accountability, and dissatisfaction with traditional tests have encouraged teachers and entire states to embrace alternative forms of assessment. Aschbacher (1991) reported that about half of the 50 states in a 1990 survey conducted by the Center for Research on Evaluation, Standards, and Student Testing (CREST) were involved to varying degrees in innovative performance assessments.

According to Stiggins (1991), large scale assessments currently account for only a small fraction of one percent of all assessment events in America's schools. The other ninety-nine percent of assessments are conducted by

teachers in classrroms day to day. Additionally, Stiggins and Bridgeford (l985) documented that 78% of teachers surveyed reported some use of structured performance tests. This survey provides evidence that the majority of classroom assessments are performance assessments, replete with the kinds of quality control problems expected when teachers have been provided virtually no assessment training. Thos that use performance assessments must be capable of ensuring the quality of the assessments used in classrooms. There is a pressing need to meet the challenge for sound assessment at the place in the educational process where teachers teach and students learn--at the classroom level.

While a wealth of research exists about the quality of classroom teacher's traditional assessments (Fleming & Chambers, l983; Carter, l984; Stiggins & Bridgeford, l985; Oescher & Kirby, l990), little research has been conducted on the quality of classroom performance assessments. Performance assessment techniques must be able to stand up to the same level of criticism given to traditional tests. For teachers and administrators, these assessments must be professionally credible, publicly accetable, and legally defensible.

Background

Dissatisfaction with the present educational system is a daily news item. Today's schools are charged with delivering a high quality education to all students in an effort to guarantee the rewards of successful learning and adulthood employment for each of their students. To complicate this challenge, society has changed dramatically; students in our schools have come from diverse backgrounds, diverse family patterns, and speak diverse native languages. There has been little change in the way of educating our students in response.

School systems are restructuring, hoping to find new solutions to the current difficulties encountered as they address the needs presented by a diverse student population and the demands of society. The appeal for restructuring has been heard in St. Charles Parish. Former Superintendent Thomas Tocco signed a joint agreement with Union Carbide to work toward restructuring our educational system. The partnership's expectations for change address broad and significant increases in student achievement, reduction of the high school drop-out rate, high skill-level placement of vocational students, raised teacher morale, and greater parental support. Included in the twelve components of the restructuring effort in St. Charles Parish are three that are significant to this study:

1. Virtually all students can learn at high levels and can be taught successfully.
2. Schools must be performance or outcome-based.
3. Assessment strategies must change.

These components of the restructuring effort involve implementation of Outcome Based Education (OBE). OBE represents a fundamental change in the way individuals are prepared for a changing world. According to Spady (l989), schools in the OBE paradigm are outcome-defined institutions offering expanded opportunity in the process of performance credentialing. As Spady defines it, OBE means "focusing and organizing all of the school's programs and instructional efforts around the clearly defined outcomes that all students should demonstrate when they leave school." The outcomes of significance are demonstrations of what students know (knowledge), can do (competencies), and are like (orientations) that will directly affect their success in facing future challenges and opportunities. An outcome is a demonstration of learning that

occurs at the end of a learning experience. It is a result of learning and a visible, observable demonstration of three things: knowledge, combined with competence, combined with "orientations"--the attitudinal, affective, motivational, and relational elements that also make up the performance. This demonstration happens in a real, live setting influenced by the factors that make up the setting or context.

Assessment of complex outcomes identified by outcome-based districts make it clear that new approaches to assessment are needed if we are to adequately assess students' ability to meet these outcomes. St. Charles Parish has identified six exit outcomes: knowledgeable person, creative producer, collaborative contributor, critical thinker, involved citizen, and self-directed learner. Given this new direction of performance or outcome-based learning, assessment strategies must also take a new direction. Assessment cannot and should not be divorced from instruction; assessment inevitably influences what is taught. The complex performances selected by teachers as the assessment tasks should provide direct measurement of real performance on important tasks. Reformed assessment practices and increased emphasis on the development and use of performance assessments are necessary if we are to adequately assess educational outcomes which cannot be assessed through traditional formats.

Although performance-based assessments have long been used by teachers for assessing student learning, in an outcome-based education environment, they are being used more frequently to make "high-stakes" decisions. As the demands for accountability to "prove" that schools are delivering instruction that produces desired student outcomes increase, teachers and administrators will need to assure that the performance assessments being

used in their classrooms and schools are professionally credible, publicly acceptable, and legally defensible.

Differences between the proponents and opponents of OBE and performance-based assessment have sparked debates that have caused confusion and have left many educators unsure as they attempt to change the course of their classroom assessment program and ultimately their school and district assessment program. A major challenge facing educators in the system's move to OBE is redesigning student assessment and reporting programs and building teacher capacity as it relates to performance assessment.

Given the increased use of formal classroom performance assessments and their importance to the instructional process, there is a need to investigate quality issues related to these assessments. This study will explore the degree to which performance assessments developed for the classroom demonstrate principles of quality classroom assessment. The general purpose of this study is to assess the quality of performance assessments developed by classroom teachers in St. Charles Parish, a district that is implementing OBE. Two specific objectives will be addressed in the study. The first is to develop an assessment instrument that reflects current thought regarding the development of performance assessments for classroom use. The second is to assess the quality of a large sample of teacher-developed performance assessments using the instrument to determine problem areas relating to sound assessment practices. This may lead to a better understanding of the real potential of performance assessments in measuring complex outcomes in the classroom and to implications for professional staff development in the field of assessment. It is a first step in moving toward assessments which will be professionally credible, publicly accepted, and legally defensible.

Methods

Sampling

The sample for this study consists of 92 performance assessments submitted by 79 teachers in St. Charles Parish. Of the 79 teachers, approximately 61% are in elementary schools, about 18% in middle schools, and 21% in high schools. The number of performance assessments per unit or per teacher varies from one to three. These teachers are currently implementing Outcome Based Education (OBE) units as a component of a district-wide restructuring effort. All received training in OBE and must submit unit plans to the district office.

Of the 92 performance assessments, approximately 55% are used by elementary teachers, about 14% by middle school teachers, and 30% by secondary teachers. Approximately 33% focused on language arts; social studies and science accounted for 24% and 21% respectively. Math assessments comprised 12% of the sample with elective courses such as Spanish, physical education, or music accounting for the remaining 11%. A list of performance assessments is found in Appendix B.

Instrumentation

The original instrument developed for this study was designed around six domains and sixteen performance criteria which helped to define each of the domains. To establish the reliability of the instrument, six doctoral students from the University of New Orleans participated in a training session in which sound assessment practices in relation to performance assessment were discussed. The instrument was the focus of this training. Each student was given a copy of three performance assessments to score independently over a three day period. Results showed interrater reliability problems on four performance criteria. Using these results, the instrument was revised. Interrater reliability of the current

instrument is 0.95 when the two developers of the instrument used it to independently score a large sample of performance assessments.

The current instrument is designed around six domains and eighteen performance criteria, each defined with clearly differentiated four point scoring scales. The six domains are (1) Purpose, Target, and Method; (2) Articulation of Performance Criteria; (3) Setting; (4) Scoring Scale; (5) Scoring Record; and (6) General Qualities. A performance assessment scoring record was developed for recording and summarizing results of the assessment. A discussion of each domain with its unique performance criteria follows. A copy of the instrument is found in Appendix C.

Domain I. Purpose, Target, and Method. Teachers collect assessment information for a purpose, and the purpose influences what will be assessed and how the assessment will be carried out. An essential feature of a quality performance assessment is a clear purpose which identifies the decision to be made from the performance assessment. How the results produced by the assessment will be used must be clearly stated.

Three primary purposes of performance assessments are to grade, to diagnose strengths and weaknesses, and to chart student improvement over time. When the purpose is diagnosis, the details of observation and scoring are different from when the purpose is to give an overall rating of pupil performance. Therefore, it is important to know *why* a performance assessment is being conducted.

The quality of any assessment depends on the appropriateness and clarity of the achievement target to be assessed. A quality performance assessment must provide a clear and appropriate description of the educational outcome, or target, it is designed to assess. General types of educational targets include mastery of knowledge, reasoning and problem solving, skill

targets, product targets, or affective targets. To design instruction and quality assessment, *what* is to be assessed, that is, the target, must be specified.

The choice of assessment method in any classroom context is a direct function of purpose and target. Method is the way in which the teacher chooses to elicit the desired outcomes; it revolves around the question of *how* to assess. Typical assessment methods include selected response (classic, objectively scored paper and pencil test), essay (extended written response), performance assessment (based either on observations of the process while skills are being demonstrated or on the evaluation of products created), and personal communication (talking and questioning students). Selecting the method of assessment that comes closest to representing the valued outcome (target) within the resources realities of the classroom is a key to sound assessment. Very simply, different purposes and different targets require different methods of assessment. The alignment between the purpose and the target and the target and method are critically important to the assessment process.

Since quality assessment arises out of the statements of the purpose (i.e., why assess), target (i.e., what to asses), and the method (i.e., how to asses), this domain is defined around three criteria. First, the purpose of the assessment must be articulated (Criterion 1). Second, the target must be articulated and focused (Criterion 2). Third, the assessment method must be matched to the target (Criterion 3).

Domain II. Articulation of Performance Criteria. In performance assessment contexts, the target is defined in terms of the performance criteria. This domain deals with the identification of observable aspects of the student's performance or product that will be judged. A key to identifying performance criteria is to break down the overall performance or product into its essential component parts that can be observed and judged. The qualities being

evaluated by a performance assessment must be described in terms of directly observable behaviors or tangible products. Performance criteria need to be specific enough to focus the teacher and student on well-defined characteristic of the performance or product. They must be developmentally appropriate for the student and useful to both the teacher and student.

Quality assessment arises out of articulation of performance criteria. This domain consists of five criteria. First, the performance criteria are specified (Criterion I). Second, the performance criteria are expressed in terms of observable behaviors or products (Criterion 2). Third, the performance criteria are comprehensive, reflecting the essential components of the task (Criterion 3). Fourth, the performance criteria are developmentally appropriate for the student (Criterion 4). Fifth, the performance assessments are comprehensible to the teacher and student (Criterion 5).

Domain III. Setting. Depending on the nature of the performance or product, the teacher may observe behaviors as they naturally occur in the classroom or set up a specific exercise or situation in which the students must perform. Generally, the more important the decision to be made from a performance assessment, the more structured the assessment environment should be. Another consideration is whether one observation of each student's performance or product will be sufficient to gather the information needed to make the decision. Multiple observations are more desirable but sometimes are limited by the amount of time it takes to complete a single observation. Therefore, providing an appropriate setting for eliciting and judging the performance or product is crucial to quality assessment.

Regardless of whether the setting is natural or structured, quality assessment arises out a consideration of the setting in which the assessment occurs. Two criteria exist for this domain. First, the student performance relative

to the performance criteria can be demonstrated in the setting in which the assessment takes place (Criterion 1). Second, the student performance can be assessed in the setting (Criterion 2).

Domain IV. Scoring Scale. The quality of performance assessments depends heavily on the scoring procedures. The nature of the decision to be made influences the scoring system used. Scoring can be analytic or holistic. With analytical scoring, judgment is made by considering each key dimension of performance or criterion separately, thus analyzing performance in terms of each of its elements. With holistic scoring, judgment is made by considering all of the criteria simultaneously, making one overall evaluation of performance. Holistic scoring is more useful where the decision to be made is a general one. If the assessment purpose is to diagnose student difficulties or certify student mastery of each individual performance criterion, then analytic scoring with a separate score or rating on each performance criterion is appropriate.

A list of performance criteria can sometimes be written in the form of a checklist. Checklists are appropriate when the process or product can be broken into components that are judged to be present or absent. Rating scales allow the observer to judge performance along a continuum rather than as a dichotomy. Both checklists and rating scales are based upon a set of performance criteria, but a checklist gives the observer two categories for judging while a rating scale gives more than two.

Quality assessment arises out of the scale by which performance criteria are scored. There are three criteria for this domain. First, the scoring scale should represent an underlying continuum of quality relevant to the performance criteria (Criterion 1). Second, points on the continuum should be specified (Criterion 2). Third, these points should differentiate the quality of the performance (Criterion 3).

Domain V. Scoring Record. Introductory educational measurement texts typically recommend guidelines for providing feedback and documenting student performance. Techniques for documenting, summarizing, and communicating results of traditional classroom assessments are a common feature of these texts. This implies that the quality of an assessment is only as good as the scoring record and communication of the results. Assessments with high-communication value provide a record that documents student performance, a clear summary of the information, and clear communication of the results. This is a common expectation of traditional assessment methods; its importance cannot be diminished when teachers use performance assessment. The need for high-communication value of the results from performance assessments cannot be left to chance.

Maintaining a written record of student performance and managing the results are essential to quality performance assessment. There are three criteria for this domain. First, the scoring record should document the performance of students on the established performance criteria (Criterion 1). Second, the scoring record should summarize the assessment using the collected data (Criterion 2). Third, the scoring record should communicate the results of the assessment (Criterion 3).

Domain VI. General Qualities. The aim of assessing student performance is to provide students a fair opportunity to demonstrate what they have learned from the instruction provided. Planning and organizing the entire performance assessment to elicit the desired demonstration of student performance is crucial. Teacher developed performance assessments must exhibit a logical and organized format that guides communication of the teacher's expectations for student performance to all participants in the process. All documents involved in the planning, development, implementation, and

communication of the performance assessment and its results must be free of grammatical errors.

Quality assessment arises out of proper format. Two general qualities are important to any quality assessment format. First, the entire performance assessment must be organized (Criterion 1). Second, the performance assessment must use standard writing conventions (i.e., syntax, usage, capitalization, punctuation, and spelling) (Criterion 2).

Procedure

The procedure used to gather information from teacher's classroom performance assessments included several steps. As noted earlier, the district in which this study was conducted had identified OBE as a vital component of their restructuring effort. The district had initiated some in-service training in this area which included the development and implementation of performance assessments. Additionally, the district implemented the use of two OBE unit formats that provided important information relative to purpose, target, and method (see Appendix D). Note that all teachers were required to submit their OBE units to the district office to receive feedback and for the purpose of identifying model units that may be used in future workshops. This provided the opportunity to explore the quality of teacher developed classroom performance assessments.

Copies of the OBE units submitted to the district during a recent period were obtained for the purpose of assessment and providing feedback to the district. From these, ninety-two samples were scored using the instrument and scoring record developed by the researchers. During a one-week period, one researcher independently scored fifty-one samples, and a second researcher scored forty-one samples. Additionally each researcher recorded descriptive information using a design characteristic form. This information included the

grade level (elementary, middle, high school), the content area (language arts, social studies, science, math, physical education, elective), the focus of the assessment (individual or group assessment), the type of performance (process/behavior, product, or a combination of process/product), the performance task, the nature of the performance task (structured assignment or naturally occurring events), and the number of performance criteria. Using the scoring record and design characteristic form, data were translated in to electronic form for compilation as well as descriptive and statistical analyses.

Results

The results of three analyses of the data are reported. The first describes the general nature of the performance assessments in our sample (e.g., the grade levels, content areas, etc.) and the ratings of those assessments. The second presents a comparative analysis of total scores across content areas and grade levels. The third reports the scores for each of the criteria in each domain.

Descriptions of the Assessments

Descriptive Information. Descriptions of the performance assessments are discussed in the sampling section of this paper. Table 1 presents much of this information in tabular form.

Insert Table 1 about here

The overwhelming majority of the assessments assessed individual (81%) rather than group (19%) performance. Approximately 70% assessed products (e.g., maps, papers, games, etc.), 16% assessed behaviors (e.g., presentations, speeches, experiments, etc.), and 14% a combination of behaviors and products (e.g., writing a play script and then performing it). Typically about nine criteria were included on any single assessment.

Descriptive Statistics. Table 1 presents the means and standard deviations across grade levels and content areas. The typical rating was 49.63. Means ranged from a high of 60.00 for the single middle school social studies assessment to a low of 39.50 for the elementary electives. Scores were somewhat varied, with higher levels of variation in social studies and science in comparison to language arts, mathematics, and elective courses.

Comparative Analyses of the Total Scores

The analysis of total scores across grade levels and content areas was problematic given the small cell sizes for 1) middle and secondary level content areas and 2) elective courses at the elementary and middle level (see Table 1). To resolve these problems, the researchers collapsed the middle and secondary levels and eliminated the elective courses from the analyses. The first decision was made on the basis of relatively greater similarity between the assessment purposes and targets of middle and secondary schools in comparison to those of elementary schools. The second reflects a common situation given the scarcity of elementary school electives.

A factorial ANOVA was used to compare the total scores across the two levels of grade level and the four levels of content. The results presented in Table 2 indicate a significant effect for grade level ($F_{1,74}$= 6.62, p. = .01) and nonsignificant effects for either content ($F_{3,74}$ = 0.73, p. =.54) or the interaction of grade level and content ($F_{3,74}$ = 0.18, p. = .91). Given only the two levels of grade level, we conclude that a significantly higher level of quality exists for secondary rather than elementary performance assessments.

Insert Table 2 about here

Description of Problematic Criteria Ratings

While the total score represents a rating of the performance assessment in general, several disconcerting patterns became apparent when analyzing the criteria scores (see Table 3). Of the 18 criteria, eight mean scores were below 3.00, a standard that we used to loosely define the difference between acceptable and unacceptable performance. These seven criteria are Domain I, Criterion 1; Domain II, Criteria 1 and 3; Domain IV, Criteria 1, 2, and 3; and Domain V, Criteria 2 and 3.

Insert Table 3 about here

Domain I. In Domain I we consistently found the purpose of the assessment articulated poorly (Criterion 1); typically we found no statement of the purpose. In a few instances we found explicit statements such as *"Spanish I Final Exam"* or a grading scale to which the score was compared. This situation is particularly disturbing given the nature of the assessments we analyzed. In case after case we felt the targets and performance criteria would change substantively depending on whether the purpose of the assessment was diagnostic or evaluative.

Domain II. In Domain II we found the specificity (Criterion 1) and comprehensiveness (Criterion 3) of the performance criteria problematic. The typical scores for Criterion 1 were either a 2 or 3. Instances in which scores were very low are exemplified by statements such as

- *most creative*
- *most artistic*
- *most informative*
- *My drawing can be recognized for what it is supposed to be.*

- *It is neat.*
- *This is my best effort.*

These were in contrast to higher scores exemplified by statements such as

- *The same verb tense is used throughout the scrapbook*
- *Did you make regular eye contact?*
- *Did you state the purpose of the interview?*
- *Lines are memorized by all performers; and scripts are not used during performance.*
- *The presentation has pre-recorded sound.*
- *A significant conflict is clearly apparent in the short story.*
- *The Illustrated Dictionary contains both print and non-print resources.*

The typical score for Criterion 3 was a 2 or 3 also. A number of lower scores are exemplified by a task statement for students to design a survey to determine the use of recycling bins. The "Survey Rubric" is limited to the following four criteria:

- *Is your name on the survey?*
- *Did you ask your parents all of the questions on the survey?*
- *Did you ask you neighbors all of the questions on the survey?*
- *Did you record all of the answers?*

A typical high score is exemplified by a task of writing an analytical paper explaining the results from stock investments. The "Market Mania Rubric" to assess that paper included the following criteria.

- *The title page includes a title, name, date, and section number.*
- *The paper includes an introductory paragraph.*
- *The paper names the stocks and briefly describes each company.*
- *Each purchase is analyzed in a separate paragraph that includes stock performance, total gains or losses, and possible reasons for each stock's performance.*

- *The paper includes a concluding paragraph.*
- *A works cited page is included.*
- *The paper is either typed or neatly written in blue or black ink.*
- *The paper is neat. No rough edges and little or no corrections are apparent.*

A second example is for an interviewing task where the "Rubric for Interviewing" included statements such as

- *Were your questions prepared in advance?*
- *Did they require more than a yes or no response?*
- *Did you state the purpose of the interview?*
- *Did you use a clear voice?*
- *Did you listen well to the entire response?*
- *Did you make regular eye contact?*
- *Did you record the interview?*
- *Did you take good notes?*
- *Were your notes brief and not take too much time?*
- *Did you thank the interviewee for their time?*
- *Did you meet the interview deadline?*

The articulation of performance criteria is critically important to the success of a performance assessment. The results from our analysis are disturbing as they empirically validate our informal assessments that teachers do not have a clear sense of how to identify and delineate the significant components of tasks.

Domain IV. In Domain IV we found problems with the underlying continuum of quality (Criterion 1), the specificity of the quality indicators (Criterion 2), and the points on the continuum (Criterion 3). The typical score for Criterion 1 was a 2. Lower scores are exemplified by scales that identified only a

dimension of inclusion (i.e., presence) and not quality. Higher scores that fully described such quality were very rare, although some are exemplified by identifying the dimensions of performance criteria using a continuum of quality such as

- *superior, above average, average or below average, unacceptable*
- *excellence, high competence, competence, not acceptable.*

A typical score on Criterion 2 was a 1 or 2. Low scores were characterized by indicators such as

- *YES* and *NOT YET*
- *YES* and *IN PROGRESS.*

Rarely did we see a high score of 4 exemplified by the following example using a 4-point rubric for an oral presentation. The quality indicators were specified as

- *A score of 4: All equipment is set up and working at the beginning time of the presentation. All members of the group speak clearly and loudly during the presentation. The presentation lasts at least 5 minutes and not longer than 10 minutes. The presentation flows smoothly. Information is presented in an interesting manner.*
- *A score of 3: Equipment is set up and working. each member of the group speaks clearly and loudly. The presentation last 5 to 10 minutes. The presentation flows.*
- *A score of 2: Each member of the group speaks. The presentation lasts 5 to 10 minutes.*
- *A score of 1:The presentation was made.*

For Criterion 3 typical scores were a 1 or 2. Low scores were similar to those for Criterion 2. Higher scores, of which there were very few, are exemplified by the following example using a four-point rubric which differentiated the points on the continuum as follows.

- A *score of 4: Includes all print and non-print resources necessary to document facts included--more than four sources.*
- *A score of 3: Includes at least four resources, both print and non-print.*
- *A score of 2: Includes at least three resources, both print and non-print.*
- *A score of 1: Includes only print or only non-print sources or has less than three sources.*

We are quite disturbed by the consistency of low scores across all criteria in Domain IV. While some of the scales (e.g., *YES* and *NOT YET*) are philosophically aligned with OBE, we believe the issue of quality was an appropriate component of all but a few of the criteria assessed. A dichotomous "presence" scale does not reflect the dimensions of quality. Without a clear sense of what the continuum is, students cannot determine where strengths and weaknesses lie, and teachers cannot reliably assess the performance. Should these assessments be used for high stakes decisions, these weaknesses become even more serious.

Domain V. In Domain V we found problems with the summarization of data (Criterion 2) and communication of results (Criterion 3) on the scoring record. The typical assessment contained some type of scoring rubric, but data were rarely summarized nor was there any indication of how the results would be used. The lack of a summary precluded a score on Criterion 2 other than a 1. In a few instances summaries such as a total score were included, and in a very few cases these summaries reflected a weighted average across criteria. The lack of summarization made any communication of results difficult. Typically the assessments required the teacher or student to draw inferences about the pattern of responses depicted on the scoring rubric. These assessments received a score of 2 on Criterion 3.

The lack of summarization of data and communicating the results through this summary are problematic. While these concerns again may be a result of the philosophical orientation of OBE, we recognize the need for a manageable record of student performance. The format of this record is closely related to the purpose of the assessment and must reflect the needs of both the teacher and student. Should these assessments be used for high stakes decisions, this documentation is of far greater importance.

Conclusion

This study was designed to generate a picture of the quality of teacher's classroom performance assessments and of the strengths and weaknesses of these assessments. We accomplished this by developing an instrument to focus on criteria that define sound assessment practices and by using this instrument to assess a sample of teacher developed classroom performance assessments in a district currently implementing OBE.

Because classroom assessments of student performance are vitally important to the teaching/learning process, it is imperative that they be of the highest possible quality. Results of the analyses presented here suggests that teacher's development of classroom performance assessments may not be as sound as they could be. Teachers trained in the use of traditional assessment methodology face real difficulties when asked to assess students using performance assessments. Problems exist in the areas of defining purpose and target and matching the method to the target, the articulation of the performance criteria, the scoring scale, and the scoring record.

The importance of this study lies in its attempt to develop a reliable scoring instrument which researchers and teachers can use to improve the quality of their performance assessments. Results reported here suggest that the utility of this instrument lies in its applicability to diagnose strengths and

weaknesses of teacher's classroom performance assessments. It remains to be seen whether the experience gained from this development effort can readily transfer to teacher's use when developing performance assessments for the classroom.

References

Aschbacher, P. R. (1991). Performance assessment: State activity, interest, and concerns. Applied Measurement in Education, 4 (4), 275-288.

Carter, Kathy. (1984). Do teachers understand principles for writing tests? Journal of Teacher Education, 35 (6), 57-60.

Fleming, M. & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. New Directions for Testing and Measurement, 19 , 29-38.

Oescher, J. & Kirby, P. (1990). Assessing teacher-made tests in secondary math and science classrooms. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Boston, MA.

Spady, W. (1989). It's time to take a close look at outcome-based education. Draft for publication in NASSP Communique.

Stiggins, R.J. (1991). Facing the challenges of a new era of educational assessment. Applied Measurement in Education, 4 (4), 263-273.

Stiggins, R.J. & Bridgeford, N.J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.

Appendix A
Review of Literature

There are three sections to this literature review. The first section discusses the nature of performance assessments. The second reviews the growth of the use of performance assessments, and the third discusses important issues relating to performance assessments.

Nature and Definition of Performance Assessments

In the current literature on performance assessment methodology, terms such as "performance assessment," "authentic assessment," "portfolio assessment," "alternative assessment," and "direct assessment" among others are used when referring to assessment differing from traditional forms. The most commonly used terms are "performance assessments" and "authentic assessments" with variations in definitions and relations between the two terms causing some confusion. Authors have suggested that the two are synonymous (Shepard in Kirst, 1991), or that performance assessment is a subcategory of authentic assessment, or that authentic assessment is a subcategory of performance assessment (Meyer, 1992; Oosterhof, 1994).

The term "alternative assessment," popularized by Wiggins (1989), conveys the idea that assessments should engage students in applying knowledge and skills in the same way they are used in the "real world" outside of school. In "What's the Difference between 'Authentic' and 'Performance' Assessment?", Meyer (1992) uses two direct writing assignments to show that performance assessment denotes the kind of student response to be examined, whereas authentic assessment denotes assessment context. Her definitions of the two terms clarify the distinction between authentic and performance assessment. In a performance assessment, the student completes or demonstrates the same behavior that the assessor desires to measure, while in

authentic assessment, the student not only completes or demonstrates the desired behavior, but also does it in a real-life context. According to Meyer (1992), it is possible for a performance assessment not to be authentic, but it is unlikely that an authentic assessment would not be a performance assessment also. Oosterhof (1994, p. 255) states that "all authentic assessments are performance assessments, but the inverse is not true." Mitchell (1992) also suggests that performance assessment is a broad term, encompassing many of the characteristics of both authentic assessment and alternative assessment. An authentic assessment involves a real application of a skill beyond its instructional context.

In the 1992 report on testing in American schools, the Office of Technology Assessment (OTA) defines performance assessment as "testing that requires a student to create an answer or a product that demonstrates his or her knowledge or skills." Performance assessment is described as a continuum of formats ranging from simple student-constructed responses to complex, comprehensive demonstrations or collections of large bodies of work over time. Feuer & Fulton (1993, p. 478) acknowledge that performance assessment is a broad term that "covers many different types of testing methods that require students to demonstrate their competencies or knowledge by creating an answer or product." They go on to describe seven common forms of performance assessments: constructed-response items, writing, oral discourse, exhibitions, experiments, and portfolios.

Airasian (1994, p. 426) defines performance assessment as "observing and judging a pupil's skill in actually carrying out a physical activity (e.g., giving a speech) or producing a product (e.g., building a birdhouse)." Stiggins (1994) states

> Performance assessments involve students in activities

> that require demonstration of certain skills and/or the creation of specified products. As a result, this assessment methodology permits us to tap many of the complex educational outcomes we value that cannot be translated into paper and pencil tests. With performance assessments we observe students while they are performing or we examine the products created, and we judge the level of proficiency demonstrated (p. 160).

For the purpose of this study, the definition and descriptions given by Airasian (1994) and Stiggins (1994) will be used. Performance assessments can be based on observations of the process while skills are being demonstrated, or on the evaluation of products being created. Evidence of achievement is in the doing and/or in the product. The index of achievement typically is a performance rating or rubric that reflects the levels of quality in the performance.

Stiggins (1987, 1994) and Airaisian (1994) agree that the purpose of performance assessment is to assess a student's ability to translate knowledge and understanding into action and that the student's response is to plan, construct, and deliver an original response. They also agree that the major advantage is the evidence of performance skills and believe that emphasizing the use of available skill and knowledge in relevant problem contexts is how performance assessment influences learning.

Formal performance assessments are those where the teacher structures the conditions in which the performance occurs and is judged. The teacher plans in advance for the behavior to occur and/or the product to be created.

Given this distinction, formal performance assessments have four distinguishing characteristics (Stiggins, Backland, & Bridgeford, 1985 in Airasian, 1994, p. 230):

1. Pupils are asked to demonstrate a process they have been taught.
2. The process to be demonstrated can be broken down into smaller steps.
3. The process to be demonstrated is directly observable.
4. Performance is judged according to performance on the smaller steps.

Growth of the Use of Performance Assessments

A great deal of attention has focused on performance assessments in the past few years. Several reasons for the current growth of interest in performance assessments can be identified in the literature. Stiggins (1991, 1994) acknowledges that performance assessment is not new methodology; it has re-emerged in recent years in response to changes in the purposes of schooling and the demand for accountability. Changes in the political, social, and economic realities have caused educators to rethink the role schools should play in our society. Suffice it to say that educators are now beginning to recognize that the old paradigm of sorting and selecting students for the social and economic system needs to give way to a new paradigm of assuring that all students attain the competencies that will permit them to be successful after their years of schooling.

Along with this change in the mission of schooling comes an increased demand for accountability. Since the 1960s, the public has become more vocal in holding schools accountable for attainment of educational results. In the 1970s, it was through behavioral objectives; in the early 1980s, emphasis was on

minimum competencies. Now in the 1990s, reference is being made to outcome-based education, which is commanding much attention in districts, schools, and classrooms across the nation. Educators have begun to define outcomes based on what students will need to know, to do, and to be like so they can contribute as productive citizens. These outcomes as valued achievement targets are complex, and few can be translated into objective paper-and-pencil, multiple choice assessments. In the article "To Use Their Minds Well: Investigating New Forms of Student Assessment," Wolf, Bixby, Glenn and Gardner (1991, p. 31) state, "There is growing, if far from universal, impatience with student assessment that addresses chiefly facts and basic skills, leaving thoughtfulness, imagination, and pursuit untapped." Therefore, the need is surfacing for a broader array of assessment techniques that includes classroom performance assessments.

A second reason for the growth of interest in performance assessment is focused on the perceived weaknesses of standardized tests. Public demand for evidence that teachers and schools are effectively educating students is increasing, and test scores are the kind of evidence the public typically finds most credible. However, there is a growing recognition that rising scores on standardized tests do not necessarily mean that students are better educated than in the past, and criticisms leveled at standardized, norm-referenced tests are widespread. Hambleton and Murphy (1992) discuss criticisms of objective tests fostering a one-right-answer mentality, narrowing the curriculum, focusing on discrete skills, and underrepresenting the performance level of low-income minority students. They argue that the evidence against multiple choice tests is not as strong as has been claimed and that more research as to the strengths and weaknesses of other assessment formats for meeting particular measurement needs should be carried out.

Worthen and Spandel (I99I) discuss the most common criticisms of standardized tests. Standardized testing is a standard operating procedure used by schools to accommodate organizational needs of accountability and may not directly promote student learning. Additionally, organizations are composed of varied individuals and interest groups whose values, beliefs, and preferences conflict. Criticism of standardized testing reflect the conflicts between various coalitions that have a stake in the use of these assessments. The first criticism is that standardized achievement tests do not promote student learning. This argument is based on the idea that achievement tests do not directly measure what goes on in the classroom. They do not enhance the learning process or provide immediate feedback needed for classroom instruction Secondly, standardized achievement and aptitude tests are poor predictors of individual students' performance. Scores on standardized tests are relatively accurate, but are limited in their ability to predict future performances of individuals. Next, the content of standardized achievement tests is often mismatched with the content emphasized in a school's curriculum and classrooms. Standardized tests are developed for broad use and attempt to sample what is typically taught to students at certain grade levels in most districts. In trying to represent everyone somewhat, standardized tests sometimes end up not representing anyone. The curriculum taught in a particular school or district may not align with what is assessed in the test. Fourth, standardized tests dictate or restrict what is taught. Since standardized test scores are used for accountability purposes, the fear is that teachers and districts will emulate the curriculum suggested by the test and neglect other important concepts. Fifth, standardized achievement and aptitude tests categorize and label students in ways that cause damage to individuals. Use of the scores from these tests to categorize students as low achievers can subject individuals to demeaning placements. Next, standardized

achievement and aptitude measures are racially, culturally, and socially biased. It is claimed that most published tests favor economically and socially advantaged children. Lastly, standardized achievement and aptitude tests measure only limited and superficial student knowledge and behaviors. The claim is that these tests measure mostly low-level, rote learning and are not able to measure higher level learning. However, Worthen and Spandel (1991) defend standardized tests as having value when used correctly but acknowledge that many tests have apparent weaknesses.

Additionally, many experts argue that performance assessments are more consistent with current theories of learning than are standardized, multiple choice assessments (Shephard, 1989). The increased acceptance of theories of learning that focus on "construction" of knowledge, emphasizing problem solving and higher order learning, and integrating affective and cognitive factors appears to strengthen the position of performance assessments. Standardized, multiple choice tests are seen by some experts as focusing on "low level," learning (Wiggins, 1992) and as emphasizing factual knowledge and " well-defined decontextualized problems" (Linn, Baker, & Dunbar, 1991), and therefore as relatively poor predictors of problem solving. However, some authors suggest that multiple choice tests can in fact assess higher order thinking skills (Mehrens, 1992), and that performance assessments sometimes test low-level, simple skills (Linn, Baker, & Dunbar, 1991). Therefore, experts in the field of assessment do not suggest that performance assessment displace traditional forms of testing, but rather that more authentic testing through the assessment of real performance adds a needed dimension to the assessment picture.

Issues Relating to Performance Assessments

There are major issues that educators must resolve if performance assessment is to reach its full potential in our schools. Worthen (1993, p. 446)

states "alternative assessment holds great promise" and should be the "backbone of assessment procedures within individual classrooms." He goes on to address twelve critical issues presenting challenges to the full potential for alternative assessments. First, he cautions that clarity of the concepts and terminology associated with alternative assessments needs to be developed into a uni-vocal language to aid the advancement of the field of study. Secondly, he cautions that there appears to be a scarcity of skeptics, those who would question and criticize alternative assessments. Self-criticism is a mechanism for continuing improvement of any movement, and there needs to be a forum for this process. In discussing the issue of support from well-informed educators, he states "the classroom teacher is the gatekeeper of effective alternative assessment (p.447)." Here he speaks to the issue of teachers' competence to perform quality assessment and refers to the report by Stiggins (1991) that suggests that educational practitioners are seriously lacking in "assessment literacy." The fourth issue is that of technical quality and truthfulness. Worthen (l993) asserts that there is little agreement about the standards that should apply to performance assessments, about what the rules of evidence should be, and about the technical specifications and criteria used to judge the quality of the assessments. Questions of validity, reliability, and generalizability among others are raised. He asserts that the "crux of the matter is whether or not the alternative assessment movement will be able to show that its assessments accurately reflect a student's true ability in significant areas of behavior that are relevant to adult life (p. 448).". The issue of standardization of assessment judgments raises the concern about how to standardize criteria and performance levels sufficiently to support necessary comparisons without causing them to lose power and richness. Especially if performance assessments are used to

inform high-stakes decisions, the question of how much standardization to introduce in alternative assessments is a key issue.

The sixth issue is one of the ability to assess complex thinking skills. The limitations of traditional measures in assessing thinking skills is a key reason for the increased use of performance assessments. However, the assumption that students are using higher-order skills whenever they are performing a hands-on task needs to be closely examined, and care needs to be taken to select assessment tasks that require students to use and demonstrate complex thinking skills. The extent to which performance assessments are acceptable to education's key stakeholders is crucial. Currently the public demand for evidence that teachers and schools are effective is answered with test scores on various measures of achievement. Proponents of the new assessments must find ways to convince stakeholders that alternative assessment can play a pivotal role in improving teaching and learning that will have benefits for the long term. There is also danger that supporters of alternative assessment may raise stakeholders' expectations to unrealistic levels and overpromise on what it can deliver. The question of alternative assessments' appropriateness for high-stakes assessment revolves around issues of standardization, bias for ethic minorities, and validity. Feasibility is the ninth issue raised revolving around issues of cost, efficiency, and the labor-intensity of developing, using, and scoring the assessments. More research on the costs verses benefits of alternative assessment needs to be conducted. When discussing the issue of continuity and integration across educational systems, Worthen (l993) suggests that the development of strategies to link assessment for accountability more effectively to assessment for individual student diagnosis and prescription must be accomplished. Use of technology to make alternative assessment less labor intensive is an important issue to be resolved. Finally, the issue of avoidance of

monopolies addressed the how to capitalize on the considerable expertise of testing corporations without abandoning to them all of the responsibility for developing local assessments. Despite his concern that there is much work ahead, Worthen (1993) advises schools to capitalize on alternative assessment whenever appropriate, because he believes it offers much at the local level.

In their argument for quality control in the development and use of performance assessments, Dunbar, Koretz, and Hoover (1991, p. 301) state "Quality control in terms of both evidence and consequences is not a question of faith, but an empirical matter when measurement is intended to inform public policy." Quellmaltz (1991, p. 319) states that the "greatest challenge facing proponents of performance assessments is the development of evaluative criteria that represent clear, significant, useful levels of expertise." She suggests that performance assessments used at either the classroom level or at a larger system level should apply quality standards that represent the consensus of professionals in the field and within the system applying the standards. Issues related to the technical standards of validity and reliability must be addressed for educators to develop useful and sound criteria. She discusses six characteristics that the criteria used to evaluate performance should possess.

1. Significance. Criteria specify important performance components; criteria specify major developmental milestones in the target domain.
2. Fidelity. Criteria represent standards that would apply appropriately within the contexts and under the conditions within which the performance typically occurs.
3. Generalizability. Criteria apply to a class or type of parallel tasks, contexts, and conditions; experienced raters

> apply the criteria consistently within and between tasks.
>
> 4. Developmental appropriateness. Criteria specify a range of quality levels appropriate for the examinee population, yet are anchored within a full, defined continuum of expertise development.
>
> 5. Accessibility. Criteria communicate clearly to and can be used by participants in the performance assessment process, including teachers, students, parents, and community.
>
> 6. Utility. Criteria communicate information about performance quality with clear implications for decision making and improvement. (p. 320)

Additionally, Quellmalz (1991) suggests some tactics for specifying criteria. Among these are surveying professional literature and seeking expert advice in the academic and practical domains, reviewing previously completed assessments, analyzing actual samples of student work and performance samples, balancing the advanced and basic skills referenced in the criteria, keeping the number of criteria manageable, and periodically reexamining the criteria to refine their understanding.

There is general agreement that validity is the most important and comprehensive concept in applying measurement standards. Messick (1994) argues that

> performance assessments must be evaluated by the same validity criteria, both evidential and consequential, as are other assessments. Indeed, such basic assessment issues

> as validity, reliability, comparability, and fairness need to be uniformly addressed for all assessments because they are not just measurement principles, they are social values that have meaning and force outside of measurement wherever evaluative judgment and decisions are made. (p. 13)

As educators increase the stakes attached to classroom performance assessments and begin to use these assessments to inform public policy, questions of quality control need to be studied empirically. Stiggins (l994) sets forth a set of guiding principles for high-quality classroom assessment that relate to target, purpose, method, sample, and control of interference. He asserts that sound assessments arise from clear achievement targets. Asking the question, "Is the target clear and appropriate?" is important here. Secondly, sound assessments arise from a clear statement of the purpose for the assessment. A third criteria for sound assessments is that the assessment must match the target and purpose of the learning outcomes. A fourth principle deals with whether sampling is appropriate given the performance target, purpose, and method. Finally, performance assessments must be designed to control for all sources of extraneous interference that can cause mismeasurement.

Resources

Aschbacher, P. R. (1991). Performance assessment: State activity, interest, and concerns. Applied Measurement in Education, 4 (4), 275-288.

Airasian, Peter W. (1994). Classroom Assessment (2nd Ed.). New York: McGraw-Hill, Inc.

Dunbar, S. B., Koretz, D.M., & Hoover, H.D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4 (4), 289-303.

Feuer, M.J. & K. Fulton (1993). The many faces of performance assessment. Phi Delta Kappan. 74(6). p. 478.

Hambleton, R. K. & Murphy, E. (1992). A psychometric perspective on authentic measurement. Applied Measurement in Education, 5 (1), 1-6

Kirst, M.W. (1991). Interview on assessment issues with Lorrie Shepard. Educational Researcher. 20 (2). 21-23, 27.

Linn, R., Barker, E., & Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, Nov., 1991, 15-21.

Meyer, Carol A. (1992). What's the Difference between authentic and performance assessment? Educational Leadership. 49 (8), 39-40.

Oosterhof, A. (1994). Classroom applications of educational measurement (2nd Ed). New York: Macmillan Publishing Company.

Quellmalz, E.S. (1991). Developing criteria for performance assessments: The missing link. Applied Measurement in Education. 4 (4), 319-331.

Shepard, Lorrie A. (1989). Why we need better assessments. Educational Leadership, 46, 4-9.

Spady, W. (1989). It's time to take a close look at outcome-based education. Draft for publication in NASSP Communique.

Stiggins, Richard J. (1987). Design and development of performance assessments. Educational Measurement: Issues and Practice, 6 (3), 33-42.

Stiggins, R. J. (1991). Facing the challenges of a new era of educational assessment. Applied Measurement in Education, 4 (4), 263-273.

Stiggins, Richard J. (1994). Student-Centered Classroom Assessment. New York: Macnillan College Publishing Co., Inc.

U.S. Congress. Office of Technology Assessment (February, 1992). Testing in American Schools: Asking the Right Questions. OTA-SET-519. Washington, D.C.: U.S. Government Printing Office. (ED 340770)

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan. 70, 703-713.

Wiggins, G. (1989). The futility of trying to teach everything of importance. Educational Leadership, 11, 44-59.

Wiggins, G. (1992). Creating tests worth taking. Educational Leadership, *49*, 26-33.

Wolf, D., J. Bixby, Glenn,J., & Gardner. H. (1991). To use their minds well: Investigating new forms of student assessment. In Gerald Grant (Ed.), Review of research in education (pp. 31-74). Washington DC: American Educational Research Association.

Worthen, B.R. (1993). Critical issues that will determine the future of alternative assessment. Phi Delta Kappan, 74 (6), 444-454.

Worthen, B.R. and Spandel, V. (1991). Putting the standardized test debate in perspective. Educational Leadership, (Feb.), 65-69.

Appendix B
Performance Assessment Samples

Sample #	Level	# of Criteria	Title
H01	E	9	Writing Checklist (Lang. Arts)
H02	E	8	Group Process Rubric (Lang. Arts)
H03	E	10	Quality Rubric Jungle Jamboree(Science)
H04	E	14	Story Rubric (Lang. Arts)
H05	E	5	Survey Rubric (Science)
H06	E	4	Poster Rubric (Science)
H07	M	6	Commercial (Lang. Arts)
H08	E	6	Presenting Interview to Class (Soc. St.)
H09	E	5	Sea Mural Rubric (Science)
H10	E	4	Oral Weather Report (Science)
H11	E	12	Letter Writing Checklist (Lang. Arts)
H12	E	7	Notecard Rubric (Soc. St.)
H13	E	12	Scrapbook Rubric (Soc. St.)
H14	E	5	Float Entries Rubric (Soc. St.)
H15	E	6	Stock Market Investing (Math)
H16	E	4	Rubric For Menu (Science)
H17	E	9	Quality Rubric for Letter Writing (Lang. Arts)
H18	E	11	Folktale Rubric (Lang. Arts)
H19	E	5	Adjusted Recipe-Quality Cupcakes (Math)
H20	E	9	Quality Newspaper Article (Lang. Arts)
H21	E	8	Quality Toy Rubric (Soc. St.)
H22	E	4	Oral Presentation of Toy (Soc. St.)
H23	E	6	Weatherman's Performance (Science)
H24	M	6	Market Mania Presentation (Math)
H25	M	8	Market Mania Paper (Math)
H26	M	14	Short Story Rubric (Lang. Arts)
H27	H	14	Annotated Bibliography Rubric (Lang. Arts)
H28	H	18	Works Consulted Paper Rubric (Lang. Arts)
H29	M	8	Interviewing Etiquette (Lang. Arts)
H30	M	18	Essay Rubric (Lang. Arts)
H31	M	7	Bilingual Recipe Book (Foreign Lang. Elective)
H32	M	11	Completing A Job Application (Lang. Arts)
H33	H	18	Owl Pellet Dissection (Science)
H34	H	10	Short Story Rubric (Lang. Arts)
H35	M	6	Factoring with Real Life Applications (Math)
H36	M	15	Scientific Method Rubric (Science)
H37	M	7	Rubric for Performance (Music Elective)
H38	H	3	Written Script Rubric (Foreign Lang. Elective)
H39	H	5	Oral Presentation (Foreign Lang. Elective)
H40	M	13	Booklet Rubric (Math)

H41	H	16	False Advertisement Report (Soc. St.)
H42	H	20	Booklet Rubric (Soc. St.)
H43	M	7	Rubric for Newspaper Article (Lang. Arts)
H44	M	4	Illustrated Dictionary (Lang. Arts)
H45	M	8	Written Research Report (Lang. Arts)
H46	H	14	Letter to Councilman (Soc. St.)
H47	M	31	Book Rubric (Lang. Arts)
H48	H	8	Newspaper Advertisement (Lang. Arts)
H49	H	16	Marketing Sales Training Manual (Elective)
H50	H	18	Resume Rubric (Elective)
H51	H	10	Speech Rubric (Lang. Arts)
O01	E	13	Rubric for Thunderstorms (Science)
O02	E	8	Rubric for Middle Ages Game (Lang. Arts)
O03	E	16	Completing Exercises (Elective)
O04	E	8	Planet Research Report (Science)
O05	E	9	Persuasive Letter on Conservation (Science)
O06	E	6	Map of School (Soc. St.)
O07	E	9	Develop Math Games (Math)
O08	E	4	Poster Rubric (Science)
O09	E	7	Louisiana Coloring Book (Soc. St.)
O10	E	5	Grandparent Booklet (Lang. Arts)
O11	E	6	Letter for Computer Donation (Math)
O12	E	8	Letter Requesting Information (Elective)
O13	E	9	Persuasive Letter Rubric (Science)
O14	E	6	Louisiana Booklet (Soc. St.)
O15	E	5	Community Awareness Rubric (Soc. St.)
O16	E	8	Computer Use Rubric (Lang. Arts)
O17	E	10	Share Picture Board Rubric (Math)
O18	E	5	Opinion Letter (Soc. St.)
O19	E	13	Christmas Gift Poster (Math)
O20	E	8	Fable Poster (Lang. Arts)
O21	E	5	Point of Light Mural Rubric (Lang. Arts)
O22	E	4	School Map (Soc. St.)
O23	E	16	Writing an Invitation Rubric (Lang. Arts)
O24	E	7	Recycling Speech (Science)
O25	E	8	Visitor Guide (Soc. St.)
O26	E	15	Nutrition Play (Elective)
O27	E	11	Family Interview (Soc. St.)
O28	E	11	Flag Mural (Soc. St.)
O29	E	8	Solar System Booklet (Science)
O30	E	5	Louisiana Plant Drawing (Science)
O31	H	4	Chemical Tests Rubric (Science)
O32	H	4	Science Project Poster (Science)
O33	H	5	Writing a Research Paper (Soc. St.)
O34	H	6	Historical Interview (Soc. St.)

O35	H	5	Historical Report (Soc. St.)
O36	H	8	Foreign Lang. Play (Foreign Lang. Elective)
O37	H	9	Multi-Media Presentation (Lang. Arts)
O38	H	5	Oral Report (Lang. Arts)
O39	H	10	Slide Show (Lang. Arts)
O40	H	9	Newscast (Soc. St.)
O41	H	13	Newscast Script (Soc. St.)

Appendix C
An Evaluation Tool for Assessment of Performance Assessments

DOMAIN I: PURPOSE, TARGET AND METHOD

Quality assessment arises out of the articulation of the purpose (i.e., why assess), target (i.e, what to assess), and method (how to assess). Three criteria exist for this domain. First, the purpose of the assessment is articulated. Second, the target is articulated. Third, the assessment method is matched to the target.

Criterion 1: The purpose of the assessment is articulated.

4 The purpose of the assessment is stated explicitly.

3 The purpose of the assessment is stated.

2 The purpose of the assessment is stated but ambiguous.

1 The purpose of the assessment is not stated or implied.

Criterion 2: The target of the assessment is articulated.

4 The target of the assessment is described explicitly at a level of specificity that clearly focuses data collection.

3 The target of the assessment is described at a level of specificity that focuses data collection.

2 The target of the assessment is described so that the focus of data collection is obscured or ambiguous.

1 The target of the assessment is not described.

Criterion 3: The assessment method is matched to the target.

4 The assessment method provides a direct view of student performance from which complete and accurate inferences from the results to the actual status of the target can be drawn.

3 The assessment method provides a direct view of student performance from which inferences from the results to the actual status of the target can be drawn.

2 The assessment method provides a direct view of student performance that does not support inferences drawn from the results to the actual status of the target.

1 The assessment method provides an indirect view of student performance.

DOMAIN II: ARTICULATION OF PERFORMANCE CRITERIA

Quality assessment arises out of the articulation of performance criteria. Five criteria exist for this domain. First, the performance criteria are specified. Second, they are expressed in terms of observable behaviors or products. Third, they reflect the breadth of performance. Fourth, they are developmentally appropriate for the student. Fifth, they are useful to the teacher and student.

Criterion 1: The performance criteria are specified.

- 4 The performance criteria are stated in a manner that clearly establishes the relevance of each of the criteria to the target.
- 3 The performance criteria are stated in a manner that establishes the relevance of most of the criteria to the target.
- 2 The performance criteria are stated in a manner that establishes the relevance of some of the criteria to the target.
- 1 The performance criteria are not stated. (A score at this level precludes scores on any other criteria in this domain.)

Criterion 2: The performance criteria are expressed in terms of observable behaviors or products.

- 4 All criteria are expressed in observable behaviors or products.
- 3 Most criteria are expressed in observable behaviors or products.
- 2 Some criteria are expressed in observable behaviors or products.
- 1 No criteria are expressed in observable behaviors or products.

Criterion 3: The performance criteria are comprehensive.

- 4 The performance criteria reflect all of the important components of performance.
- 3 The performance criteria reflect most of the important components of performance.
- 2 The performance criteria reflect some of the important components of performance.
- 1 The performance criteria do not reflect the important components of performance.

Criterion 4: The performance criteria are developmentally appropriate for the student.

- 4 All of the performance criteria are developmentally appropriate for the student.
- 3 Most of the performance criteria are developmentally appropriate for the student.
- 2 Some of the performance criteria are developmentally appropriate for the student.
- 1 None of the performance criteria are developmentally appropriate for the student.

Criterion 5: The performance criteria are useful to the teacher and student.

- 4 The performance criteria can be clearly understood and are very useful to the teacher and student.
- 3 The criteria are articulated, comprehensible and useful.
- 2 The criteria are somewhat comprehensible, and minimally useful.
- 1 The criteria are incomprehensible and of minimal value.

DOMAIN III: SETTING

Quality assessment arises out of the setting in which the assessment occurs. Regardless of whether the setting is natural or structured, two criteria exist for this domain. First, student performance relative to the performance criteria can be demonstrated in the setting in which the assessment takes place. Second, student performance can be assessed in the setting.

Criterion 1: Student performance can be demonstrated in the setting

4 The student can easily and thoroughly complete all performance criteria in the setting.

3 The student can complete all performance criteria in the setting.

2 The student can complete only some of the performance criteria because the setting interferes with the process.

1 The student cannot complete the performance criteria i the setting.

Criterion 2: Student performance can be assessed in the setting

4 All performance criteria can be accurately and reliably assessed.

3 All performance criteria can be assessed, but the accuracy and/or reliability of the assessment is impaired by the setting.

2 The setting hinders the assessment of some of the performance criteria.

1 The setting severely impedes the assessment of performance criteria.

DOMAIN IV: SCORING SCALE

Quality assessment arises out of the scale by which performance criteria are scored. Three criteria exist for this domain. First, the scoring scale represents an underlying continuum of quality relevant to the performance criteria. Second, points on the continuum are specified. Third, these points differentiate the quality of the performance.

Criterion 1: The scoring scale represents an underlying continuum of quality

4 The scoring scale identifies a continuum of quality that is highly relevant to the dimensions of the performance criteria.

3 The scoring scale identifies a continuum of quality that is relevant to the dimensions of the performance criteria.

2 The scoring scale identifies a continuum of quality that is remotely relevant to the dimensions of the performance criteria.

1 The scoring scale identifies a continuum of quality that is not relevant to the dimensions of the performance criteria.

Criterion 2: The points on the continuum are specified

4 The quality indicators at each point on the continuum are precise and unique.

3 The quality indicators at each point on the continuum are specific.

2 The quality indicators at each point on the continuum are vague.

1 The quality indicators on the continuum are ambiguous.

Criterion 3: The points on the continuum differentiate the quality of performance

4 The differences between indicators describe meaningful differences in the performance criteria.

3 The differences between indicators describe somewhat meaningful differences in the performance criteria.

2 The differences between indicators simply describe differences in the performance criteria.

1 The differences between the indicators do not describe differences in the performance criteria.

DOMAIN V: SCORING RECORD

Quality assessment arises out of a written record for scoring and managing results. Three criteria exist for this domain. First, the scoring record documents the performance of students on the performance criteria. Second, it summarizes the assessment using collected data. Third, it communicates the results of the assessment.

Criterion 1: The scoring record documents the performance of students

4 The scoring record provides complete and accurate information for all performance criteria.

3 The scoring record provides complete information for most of the performance criteria.

2 The scoring record provides information for few of the performance criteria.

1 The scoring record does not exist, or it provides incomplete or inaccurate information for all of the performance criteria.

Criterion 2: The scoring record summarizes the assessment using collected data

4 All information is summarized in a clear, concise manner.

3 All information is summarized, but the summary is somewhat confusing.

2 Some of the information is summarized, but the summary is confusing and imprecise.

1 Most of the information is not summarized.

Criterion 3: The scoring record communicates the results

4 The results can be easily interpreted.

3 The results can be interpreted.

2 The results are difficult to interpret.

1 The results cannot be interpreted.

DOMAIN VI: GENERAL QUALITIES

Quality assessment arises out of format. Two criteria exist for this domain. First, the entire performance assessment is organized. Second, it uses standard writing conventions (i.e., syntax, usage, capitalization, punctuation, and spelling).

Criterion 1: The entire performance assessment is organized

4 The performance assessment is clearly and logically organized.

3 The performance assessment is organized.

2 The performance assessment is somewhat disorganized.

1 The performance assessment is disorganized.

Criterion 2: The performance assessment uses standard writing conventions

4 Standard writing conventions are followed, and the document is free of errors.

3 Standard writing conventions are followed. Some errors are present, but these are few and minor in nature.

2 Standard writing conventions are followed, but numerous errors are present. These errors do not block meaning, but they impair readability and use of the instrument.

1 Numerous writing errors are present. The frequency and severity of errors make it difficult or impossible to read or use the instrument.

PERFORMANCE ASSESSMENT SCORING RECORD

Name_____ Performance Assessment Number_____

Date_____ Total Score_____

Directions: Use the attached descriptions of the performance criteria and scoring rubrics to evaluate each of the six (6) domains. Circle your score for each of the criteria. Sum all scores and record the result in the blank for total points. You can make comments on any criteria in the spaces provided.

DOMAIN	PERFORMANCE CRITERIA	SCORE			
I	Criterion 1: Purpose is articulated	4	3	2	1
	Criterion 2: Target is focused	4	3	2	1
	Criterion 3: Method is matched to target	4	3	2	1
II	Criterion 1: Performance criteria are specified	4	3	2	1
	Criterion 2: Performance criteria are observable	4	3	2	1
	Criterion 3: Performance criteria are comprehensive	4	3	2	1
	Criterion 4: Performance criteria are developmentally appropriate	4	3	2	1
	Criterion 5: Performance criteria are comprehensible	4	3	2	1
III	Criterion 1: Performance can be demonstrated in the setting	4	3	2	1
	Criterion 2: Performance can be assessed in the setting	4	3	2	1
IV	Criterion 1: Scoring scale represents an underlying continuum	4	3	2	1
	Criterion 2: Quality indicators are specified	4	3	2	1
	Criterion 3: Points on the continuum differentiate quality	4	3	2	1
V	Criterion 1: Scoring record documents performance	4	3	2	1
	Criterion 2: Scoring record summarizes assessment	4	3	2	1
	Criterion 3: Scoring record communicates results	4	3	2	1
VI	Criterion 1: Performance assessment is organized	4	3	2	1
	Criterion 2: Uses standard writing conventions	4	3	2	1

Comments

Appendix D
OBE Unit Formats

ST. CHARLES PARISH PUBLIC SCHOOLS

Unit Planner for Outcome-Based Education

Teacher _____ School _____ Title of Unit _____

Implementation Date _____

(Circle the major learner outcomes developed through this unit.)

KNOWLEDGEABLE COMPETENT PERSON	COLLABORATIVE CONTRIBUTOR	INVOLVED CITIZEN	CREATIVE PRODUCER	SELF-DIRECTED ACHIEVER	CRITICAL THINKER

Sphere(s) Of Living

(Check those that apply to this unit.)

- [ ] Personal
- [ ] Learning
- [ ] Work
- [ ] Relationships
- [ ] Cultural
- [ ] Global
- [ ] Civic

Life Issue/Question Or Significant Concept

Fundamental Life Performance(s)

(Check those that apply to this unit.)

- [ ] Learner
- [ ] Communicator
- [ ] Thinker
- [ ] Team member & peer
- [ ] Teacher & mentor
- [ ] Leader and Organizer
- [ ] Creator & producer
- [ ] User & performer
- [ ] Problem finder & solver
- [ ] Other _____

Content/Information

Enabling Processes/Competencies

Authentic Task(s): During this unit the students will ...

BEST COPY AVAILABLE

AUTHENTIC TASK DESCRIPTION

You are _____ (Role)

who _____ (Task)

_____

using _____ (Content/Process

_____ Knowledge)

in order to _____ (Performance)

_____

for _____ (Audience)

at or in _____ (Setting)

You are faced with _____ (Problems, Issues,

_____ Circumstances,

_____ Dilemmas)

You are expected to ...

1. _____ (Products/performances/ demonstrations to be completed by students.)
2. _____
3. _____
4. _____

using _____ (Support

_____ Materials)

You will be assessed according to: (List the criteria, describe or attach the rubrics, or identify the methods of assessment that will be used.)

_____

_____

_____

_____

BEST COPY AVAILABLE

ST. CHARLES PARISH PUBLIC SCHOOLS

Unit Plan for Outcome-Based Education

1. Name of Teacher and School, Title of Unit, and Implementation Date

2. Learner Outcomes Addressed
   - **Collaborative Contributor**
   - **Creative Producer**
   - **Critical Thinker**
   - **Involved Citizen**
   - **Knowledgeable Competent Person**
   - **Self-Directed Achiever**

3. Authentic Task: During this unit the student will . . .
   (Be Brief)

4. Sphere(s) of Living Addressed
   - **Personal**
   - **Learning**
   - **Civic**
   - **Work**
   - **Relationships**
   - **Cultural**
   - **Global**

5. Life Issue(s)/Question(s) or Significant Concept

6. Fundamental Life Performance(s)
   - **Learner**
   - **Communicator**
   - **Thinker**
   - **Team Member and Peer**
   - **Teacher and Mentor**
   - **Creator and Producer**
   - **Problem Finder and Solver**
   - **User and Performer**
   - **Leader and Organizer**

7. Content/Information

8. Enabling Processes/Competencies

9. Authentic Task Description and Assessment
   (Use the Authentic Task Description Form.)

AUTHENTIC TASK DESCRIPTION

You are _____ *(Role)*

who _____ (Task)

_____

using _____ (Content/Process

_____ Knowledge)

in order to _____ (Performance)

_____

for _____ (Audience)

at or in _____ (Setting)

You are faced with _____ (Problems, Issues,

_____ Circumstances,

_____ Dilemmas)

You are expected to ...

1. _____ (Products/performances/ demonstrations to be completed by students.)
2. _____
3. _____
4. _____

using _____ (Support

_____ Materials)

You will be assessed according to: (List the criteria, describe or attach the rubrics, or identify the methods of assessment that will be used.)

_____

_____

_____

_____

BEST COPY AVAILABLE

Table 1

Descriptive Statistics for Performance Assessment Ratings across Content Areas and Grade Levels

	Grade Level			
Content Area	Elementary	Middle	Secondary	Total
Language Arts				
Obs	14	6	10	30
Mean	50.21	53.17	51.80	51.33
SD	4.35	3.60	3.12	3.89
Social Studies				
Obs	13	1	8	22
Mean	48.69	60.00	51.25	50.13
SD	6.59	-	7.13	6.94
Science				
Obs	15	0	4	19
Mean	46.67	-	50.50	47.47
SD	5.89	-	7.94	6.33
Mathematics				
Obs	7	3	1	11
Mean	48.14	52.67	53.00	49.82
SD	3.34	5.13	-	4.17
Electives				
Obs	2	3	5	10
Mean	39.50	48.00	50.00	47.30
SD	2.12	5.29	2.24	5.17
Total				
Obs	51	13	28	92
Mean	48.08	52.38	51.18	49.63
SD	5.60	4.94	4.97	5.57

Table 2

Factorial ANOVA of Ratings across Grade Levels and Content Areas

Source	df	SS	MS	F
Grade Level	1	191.88	191.88	6.62*
Content Area	3	63.77	21.26	1.03
Grade*Content Interaction	3	15.91	5.30	0.18
Error	74	2144.06	28.97	

*p. = .012

Table 3

Descriptive Statistics for Criteria Scores

Domain	Criteria	Description	Mean	SD
I	1	Purpose is articulated	1.08	0.34
	2	Target is focused	3.28	0.63
	3	Method is matched to target	3.60	0.56
II	1	Criteria are specified	2.87	0.61
	2	Criteria are observable	3.55	0.73
	3	Criteria are comprehensive	2.62	0.67
	4	Criteria are developmentally appropriate	3.75	0.60
	5	Criteria are comprehensible	3.04	0.74
III	1	Can demonstrate in setting	3.65	0.56
	2	Can assess in setting	3.62	0.64
IV	1	Scale represents continuum	2.09	0.83
	2	Indicators are specified	1.85	0.75
	3	Point on the continuum differentiate quality	1.47	0.80
V	1	Record documents performance	3.48	1.01
	2	Record summarizes performance	1.17	0.55
	3	Record communicates results	1.79	0.57
VI	1	Assessment is organized	3.04	0.47
	2	Assessment uses standard writing conventions	3.67	0.49